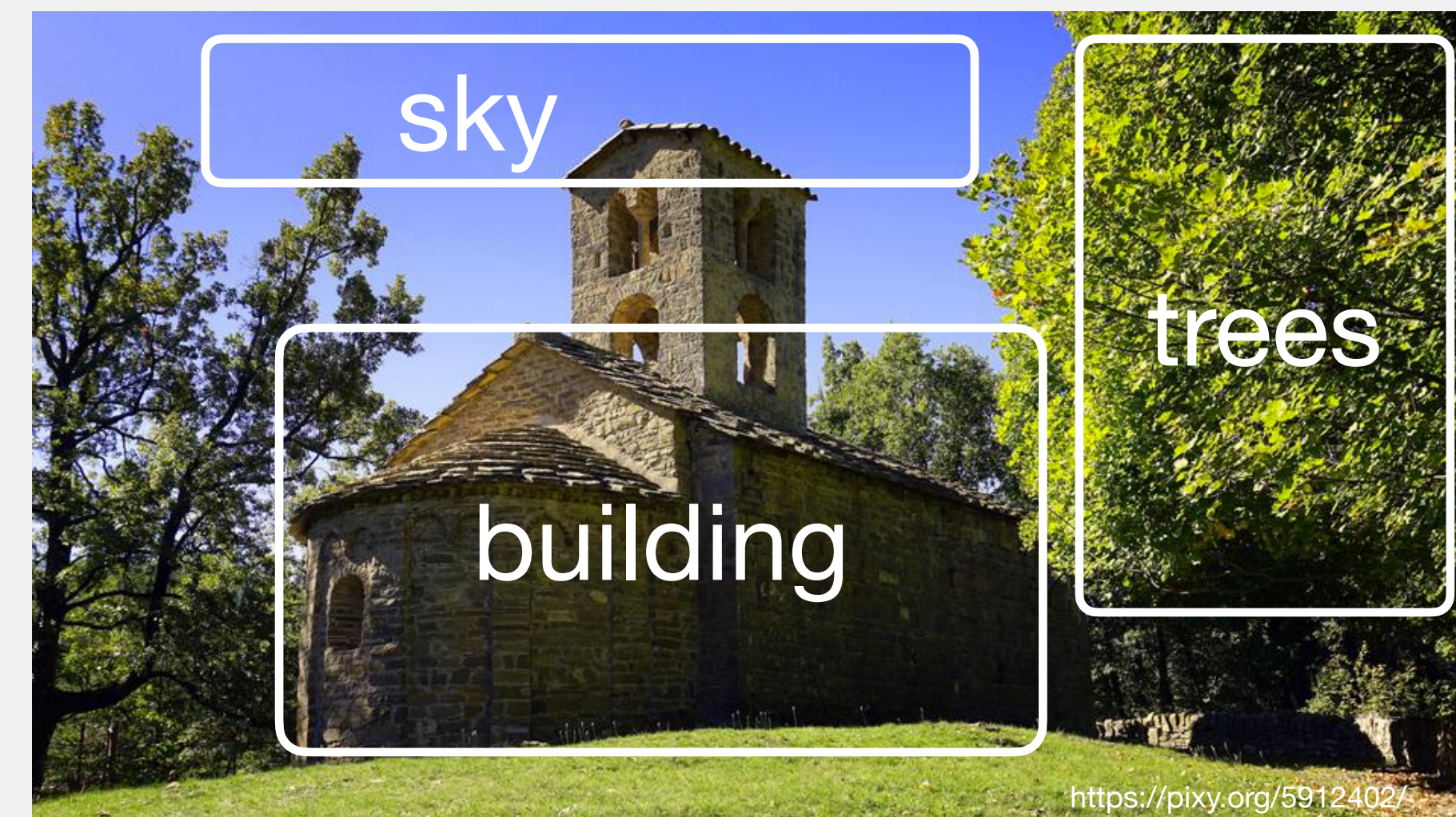


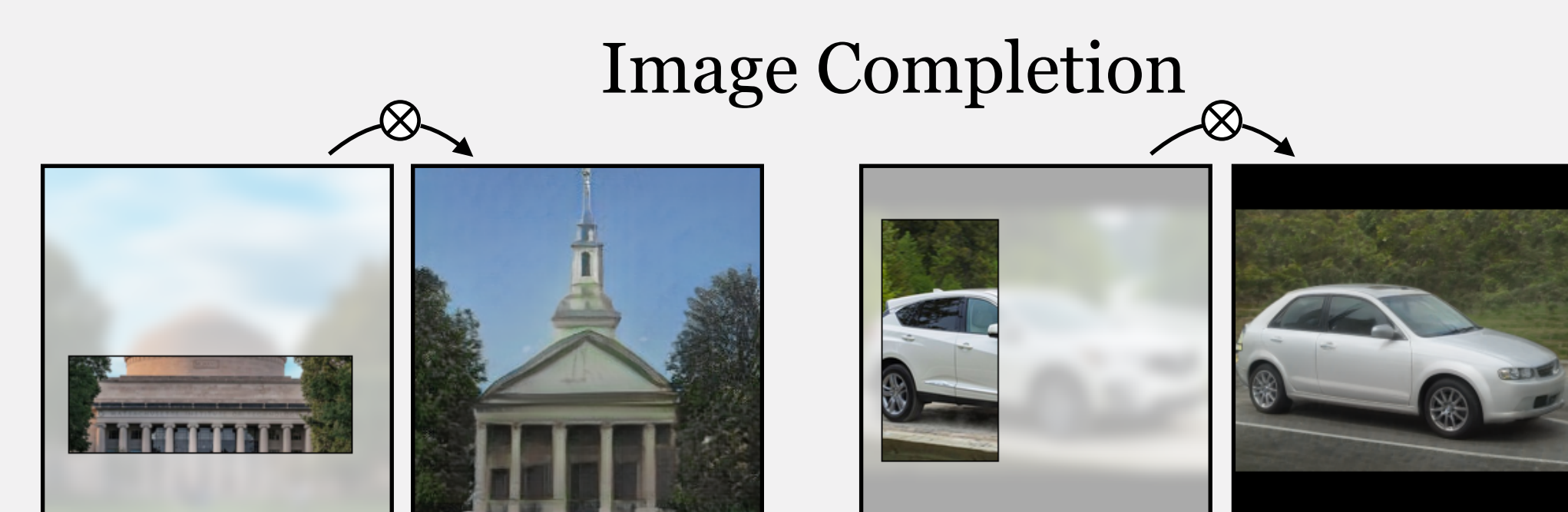
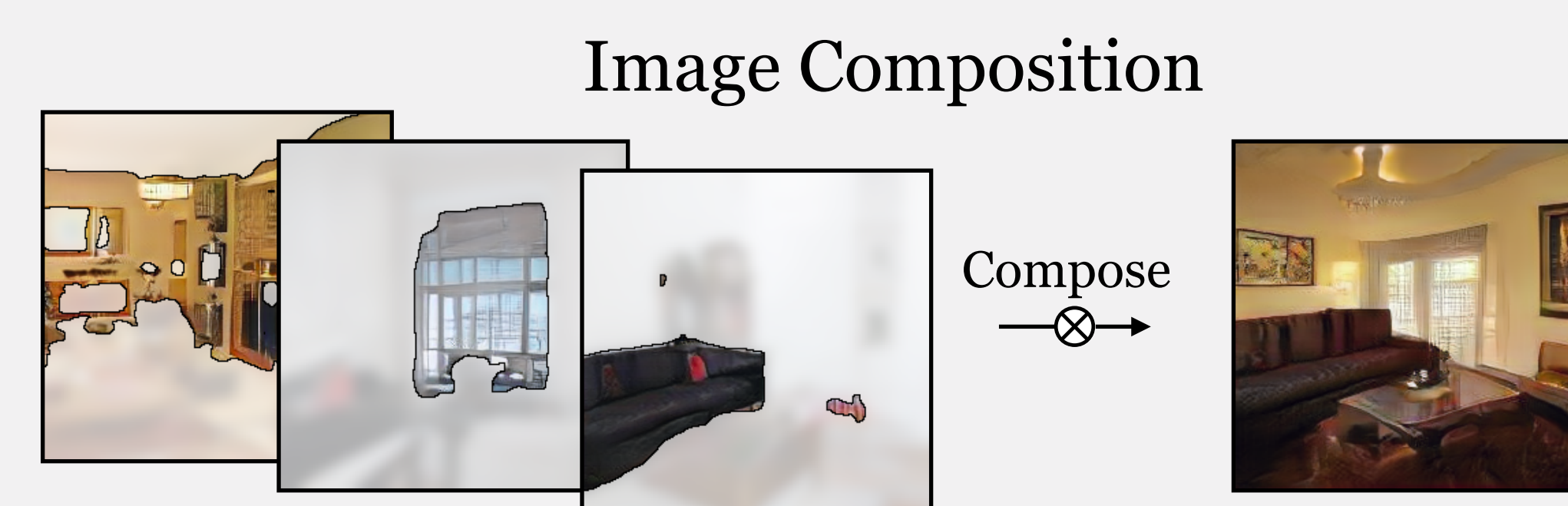
Using Latent Space Regression to Analyze and Leverage Compositionality in GANs

Lucy Chai, Jonas Wulff, Phillip Isola
{lrchai, wulff, phillipi}@mit.edu

Introduction

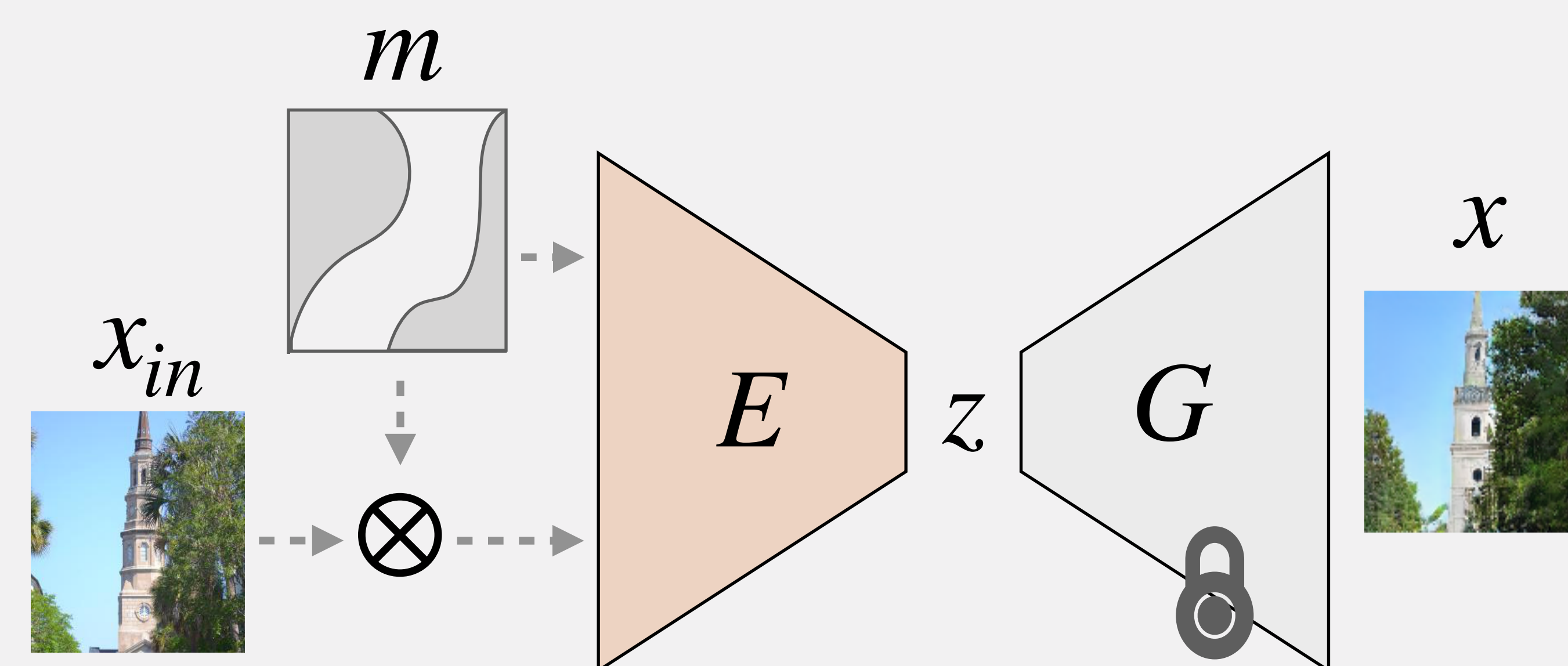


- Natural scenes are compositional, consisting of parts that can vary independently
- How do GANs represent compositional parts in latent space?



- Masked latent regression + pretrained GAN forms a strong image prior.
- Unsupervised compositional structure emerges in latent space.
- Allows real-time editing with single image examples.
- Regressor probes scene part independence.

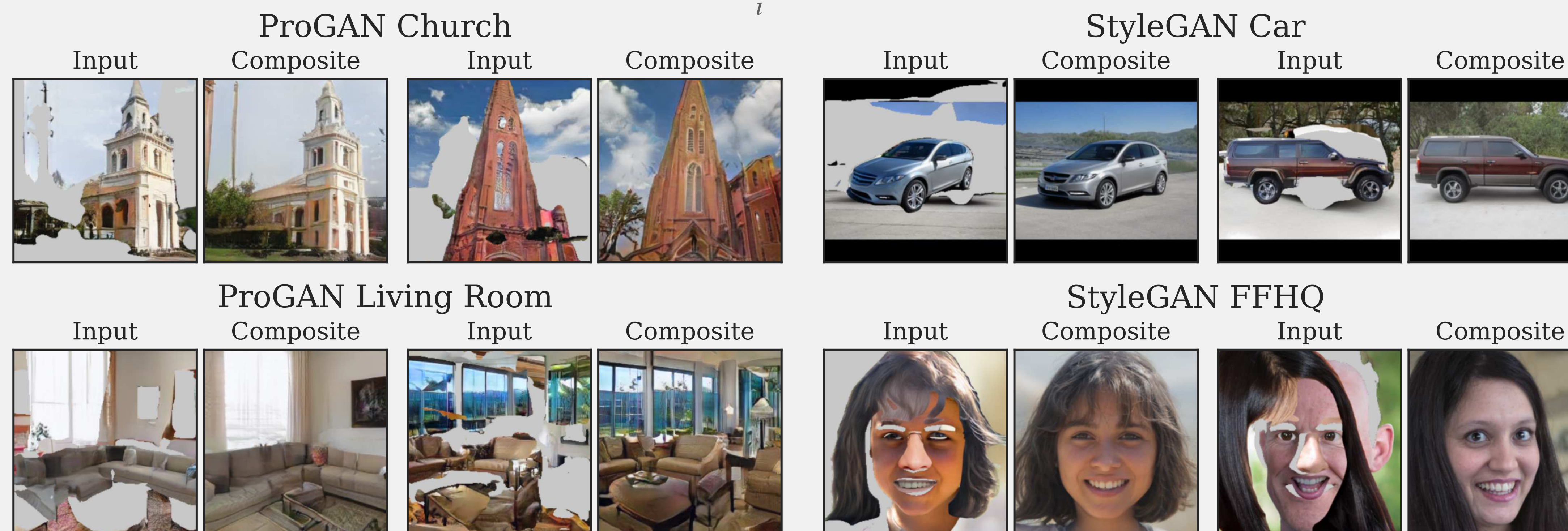
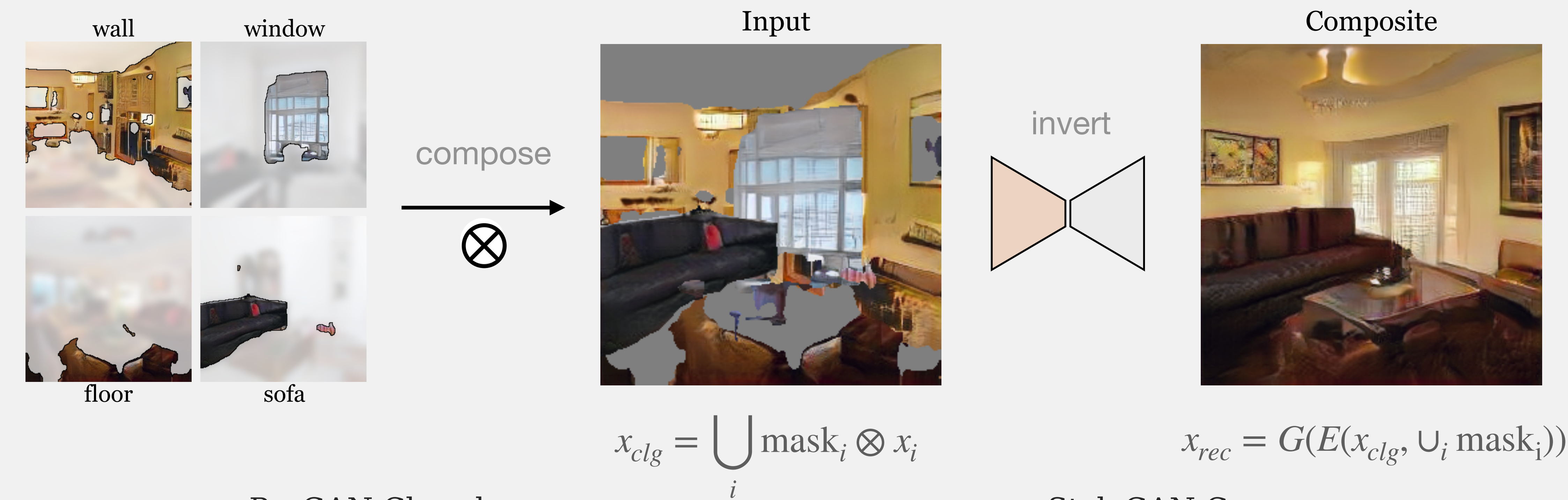
Objective Function



$$\mathbb{E}_{z \sim \mathcal{N}(0,1); x=G(z)} L_2(x, GE(x, m)) + L_p(x, GE(x, m)) + L_z(z, E(x, m))$$

- We use L2 and perceptual image losses, and a latent recovery loss.
- The added mask input specifies where the GAN must inpaint.
- This creates buffer regions to let the image prior fill in the unknown region.

Compositionality in GANs



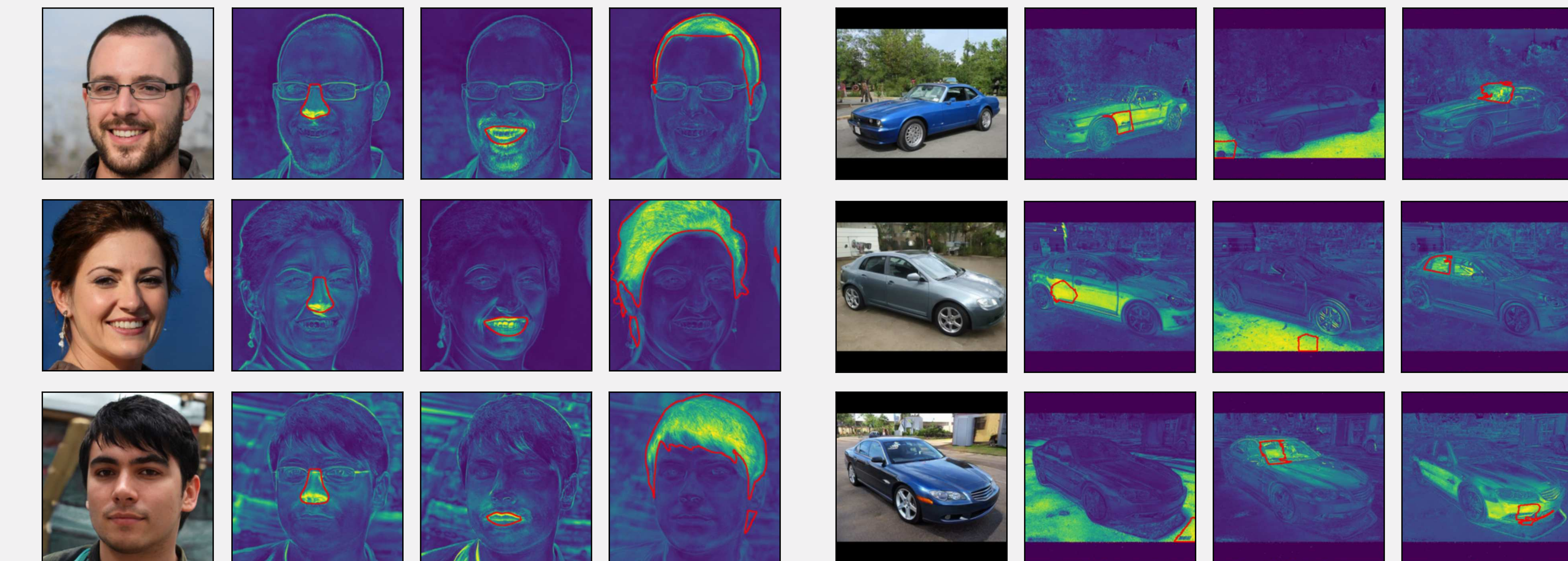
- We extract random image parts and compose them via union.
- Inversion converts the unrealistic input collage into a coherent scene.
- Creating composite images requires simultaneous blending, alignment, and inpainting, which is possible under the image prior.

Compose Operation Allows Precise Editing



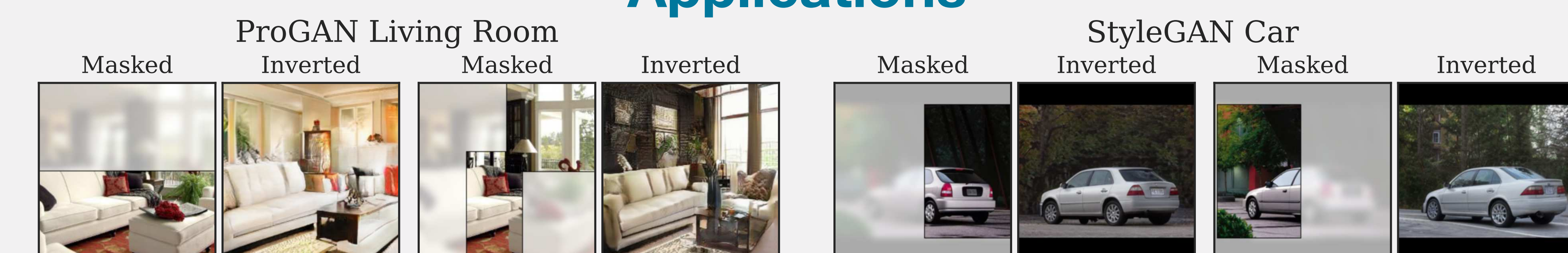
- Composition can add windows while preserving the overall scene consistency.
- Interpolation in the latent code or pixel values fails to add windows or distorts the remainder of the image.
- Composition attains lower L1 distance to the inputs compared to interpolation.

Unsupervised Discovery of Parts

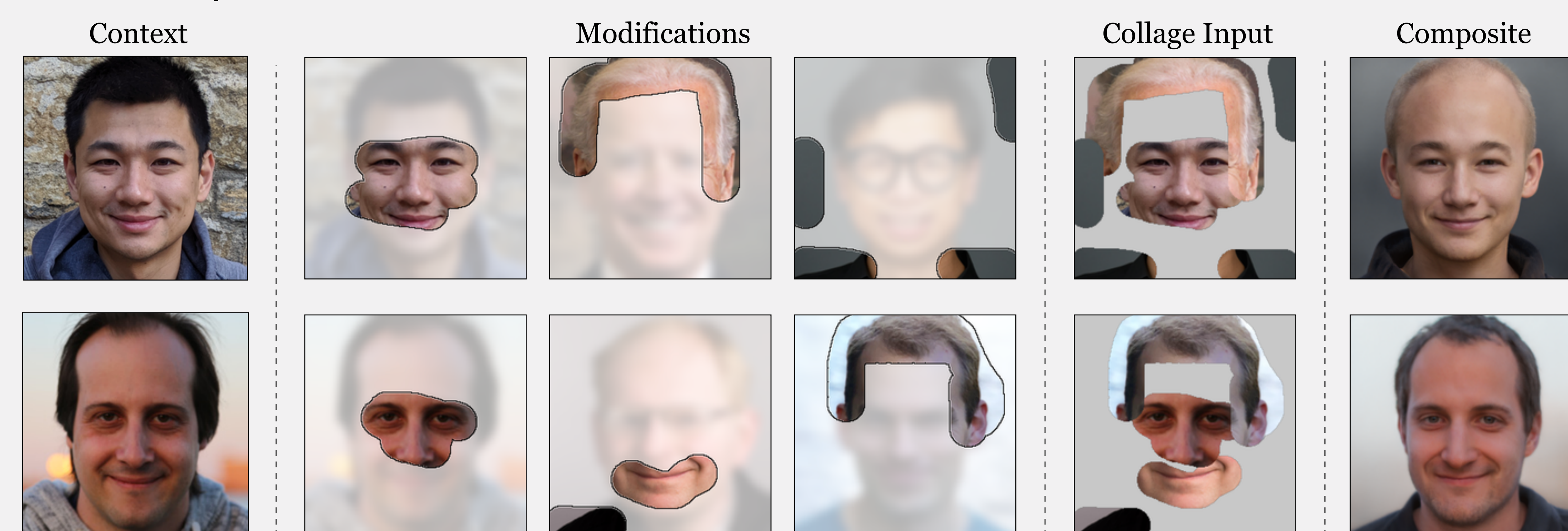


- How does the GAN understand independence of object parts in latent space?
- Visualizations show pixels that are most sensitive to change when a given patch of the image is modified.
- On faces, the variation is usually strongest in the part that is changed, matching our intuitive understanding of a face.
- Using superpixels in cars, the network learns that pixels on the same part usually change together.

Applications



- The completion of a partial image can vary based on the exposed region of the same input.



- We finetune the encoder towards a specific image. The remaining compositional edits can occur in real time.